CLAIMS

1.   A harmonic structure acoustic signal detection method of detecting a segment that includes speech, as a speech segment, from an input acoustic signal, said method comprising:

5   an acoustic feature extraction step of extracting an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period; and

a segment determination step of evaluating continuity of the acoustic features and of determining a speech segment according to
10   the evaluated continuity,

wherein in said acoustic feature extraction step, frequency transform is performed on each of the frames into which the input acoustic signal is divided at every predetermined time period, and the acoustic feature that is a value of a harmonic structure
15   represented by a number is extracted, and

in said segment determination step, the speech segment is determined based on one of the following: a correlation value between acoustic features in the same frame; and a correlation value between acoustic features in different frames.

20

2.   The harmonic structure acoustic signal detection method according to Claim 1,

wherein in said acoustic feature extraction step, a harmonic structure is further accentuated based on each component obtained
25   through the frequency transform, and the acoustic feature is extracted.

3.   The harmonic structure acoustic signal detection method according to Claim 2,
30   wherein in said acoustic feature extraction step, a harmonic structure is further extracted from each component obtained through the frequency transform, and a component which is

obtained through the frequency transform and has a predetermined frequency band that includes the harmonic structure is judged to be the acoustic feature.

4.   The harmonic structure acoustic signal detection method according to Claim 1,
    wherein in said acoustic feature extraction step, each component obtained through the frequency transform of each frame is further divided into frequency bands of a predetermined bandwidth, a correlation value is calculated between the components that have predetermined frequency bands in the same frame, and the acoustic feature is extracted based on the calculated correlation value.

5.   The harmonic structure acoustic signal detection method according to Claim 4,
    wherein in said acoustic feature extraction step, a difference is further calculated between a maximum value and a minimum value of the correlation values in each frame, and the acoustic feature is extracted based on the difference.

6.   The harmonic structure acoustic signal detection method according to Claim 1,
    wherein in said acoustic feature extraction step, each component obtained through the frequency transform of each frame is further divided into frequency bands of a predetermined bandwidth, a correlation value is calculated between the components that have predetermined frequency bands in different frames, and the acoustic feature is extracted based on the calculated correlation value.

7.   The harmonic structure acoustic signal detection method

according to Claim 6,

wherein in said acoustic feature extraction step, a difference is further calculated between a maximum value and a minimum value of the correlation values in each frame, and the acoustic
5   feature is extracted based on the difference.

8.   The harmonic structure acoustic signal detection method according to Claim 1,

wherein in said segment determination step, continuity of the
10   acoustic features is evaluated based on a correlation value between the acoustic features of different frames, and the speech segment is determined according to the evaluated continuity.

9.   The harmonic structure acoustic signal detection method
15   according to Claim 1,

wherein in said segment determination step, continuity of the acoustic features is evaluated based on distributions of the acoustic features in different frames, and the speech segment is determined according to the evaluated continuity.
20

10.   The harmonic structure acoustic signal detection method according to Claim 1, comprising:

an evaluation step of calculating an evaluation value for evaluating the continuity of the acoustic features; and
25   a speech segment determination step of evaluating temporal continuity of the evaluation values and of determining a speech segment according to the evaluated temporal continuity.

11.   The harmonic structure acoustic signal detection method
30   according to Claim 10,

wherein said segment determination step further includes:
a step of estimating a speech signal-to-noise ratio of the input

acoustic signal based on comparisons, for a predetermined number of frames, between (i) acoustic features extracted in said acoustic feature extraction step or the evaluation values calculated in said evaluation step and (ii) a first predetermined threshold; and

5   a step of determining the speech segment based on the evaluation value calculated in said evaluation step, in the case where the estimated speech signal-to-noise ratio is equal to or higher than a second predetermined threshold, and

in said speech segment determination step, the temporal

10   continuity of the evaluation values is evaluated and the speech segment is determined according to the evaluated temporal continuity, in the case where the speech signal-to-noise ratio is lower than the second predetermined threshold.

15   12.   The harmonic structure acoustic signal detection method according to Claim 1,

wherein said segment determination step includes:

an evaluation step of calculating an evaluation value for . evaluating the continuity of the acoustic features; and

20   a non-speech harmonic structure segment determination step of evaluating temporal continuity of the evaluation values and determining, according to the evaluated temporal continuity, a non-speech harmonic structure segment that has a harmonic structure but is not a speech segment.

25

13.   The harmonic structure acoustic signal detection method according to Claim 12,

wherein said acoustic feature extraction step includes:

a frequency transform step of performing frequency

30   transform on each of the frames into which the input acoustic signal is divided at every predetermined time period;

a correlation value calculation step of dividing a component

obtained through the frequency transform of each frame into frequency bands of a predetermined bandwidth, and of calculating a correlation value between the components that have predetermined frequency bands in the same frame; and

5   an extraction step of extracting, as the acoustic feature, an identifier of a frequency band in which the component has a maximum value or a minimum value of the correlation values in the same frame.

10   14.   The harmonic structure acoustic detection method according to Claim 1,

wherein said acoustic feature extraction step includes:

a frequency transform step of performing frequency transform on each of the frames into which the input acoustic signal

15   is divided at every predetermined time period;

a correlation value calculation step of calculating a correlation value between components obtained through the frequency transform of frames which are a predetermined number of frames away from each other; and

20   an acoustic feature extraction step of extracting the acoustic feature that is a value of a harmonic structure represented by a number, by calculating a distribution of the correlation values in every predetermined number of frames.

25   15.   The harmonic structure acoustic signal detection method according to Claim 1,

wherein in said segment determination step, the continuity is evaluated based on correlation values between two or more types of frames of different time periods.

30

16.   The harmonic structure acoustic signal detection method according to Claim 15,

wherein in said segment determination step, one of the correlation values between the two or more types of frames of different time periods is selected based on a speech signal-to-noise ratio of the input acoustic signal, and the continuity is evaluated based on the selected correlation value.

17. The harmonic structure acoustic signal detection method according to Claim 1,

wherein in said segment determination step, the continuity is evaluated based on a corrected correlation value calculated using a difference between (i) a correlation value between the acoustic features of frames and (ii) an average value of the correlation values of a predetermined number of frames.

18. The harmonic structure acoustic signal detection device according to Claim 28,

wherein said acoustic feature extraction unit is operable to perform frequency transform on each of frames into which the input acoustic signal is divided at every predetermined time period, and to extract the acoustic feature that is a value of a harmonic structure represented by a number, and

said segment determination unit is operable to determine the speech segment based on one of the following: a correlation value between acoustic features in the same frame; and a correlation value between acoustic features in different frames.

19. A speech recognition device which recognizes speech included in an input acoustic signal, said device comprising:

an acoustic feature extraction unit operable to extract an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period;

a segment determination unit operable to evaluate continuity

of the acoustic features, and to determine a speech segment according to the evaluated continuity; and

a recognition unit operable to recognize speech in the speech segment determined by said segment determination unit,

5 wherein said acoustic feature extraction unit is operable to perform frequency transform on each of the frames into which the input acoustic signal is divided at every predetermined time period, and to extract the acoustic feature that is a value of a harmonic structure represented by a number, and

10 said segment determination unit is operable to determine the speech segment based on one of the following: a correlation value between acoustic features in the same frame; and a correlation value between acoustic features in different frames.

15 20. A speech recording device which records speech included in an input acoustic signal, said device comprising:

an acoustic feature extraction unit operable to extract an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period;

20 a segment determination unit operable to evaluate continuity of the acoustic features, and to determine a speech segment according to the evaluated continuity; and

a recording unit operable to record the input acoustic signal in the speech segment determined by said segment determination

25 unit,

wherein said acoustic feature extraction unit is operable to perform frequency transform on each of the frames into which the input acoustic signal is divided at every predetermined time period, and to extract the acoustic feature that is a value of a harmonic

30 structure represented by a number, and

said segment determination unit is operable to determine the speech segment based on one of the following: a correlation value

between acoustic features in the same frame; and a correlation value between acoustic features in different frames.

21.　A program which causes a computer to execute:

5　　　an acoustic feature extraction step of extracting an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period; and

a segment determination step of evaluating continuity of the acoustic features and of determining a speech segment according to

10　the evaluated continuity,

wherein in said acoustic feature extraction step, frequency transform is performed on each of the frames into which the input acoustic signal is divided at every predetermined time period, and the acoustic feature that is a value of a harmonic structure

15　represented by a number is extracted, and

in said segment determination step, the speech segment is determined based on one of the following: a correlation value between acoustic features in the same frame; and a correlation value between acoustic features in different frames.

20